

Quality of the Administrative Record Data Used in the 2020 Census

National Advisory Committee
Date: May 6, 2022

Thomas Mule
Special Assistant to the Chief,
Decennial Statistical Studies Division



Data presented were approved for dissemination by the Census Bureau Disclosure Review Board
Results are covered by (CBDRB-FY20-ACSO003-B0011, CBDRB-FY21-DSSD007-0012 and CBDRB-FY22-172)

Administrative Record Data Used in the 2020 Census

Fall 2021 Recommendation:

The NAC recommends that the Census Bureau brief the NAC at its Spring 2022 meeting on the quality of the key administrative records sources used in the 2020 Census and other Census Bureau data collections, and how accurate and complete the administrative records and linkages are for historically undercounted population groups, such as persons of color and young children.

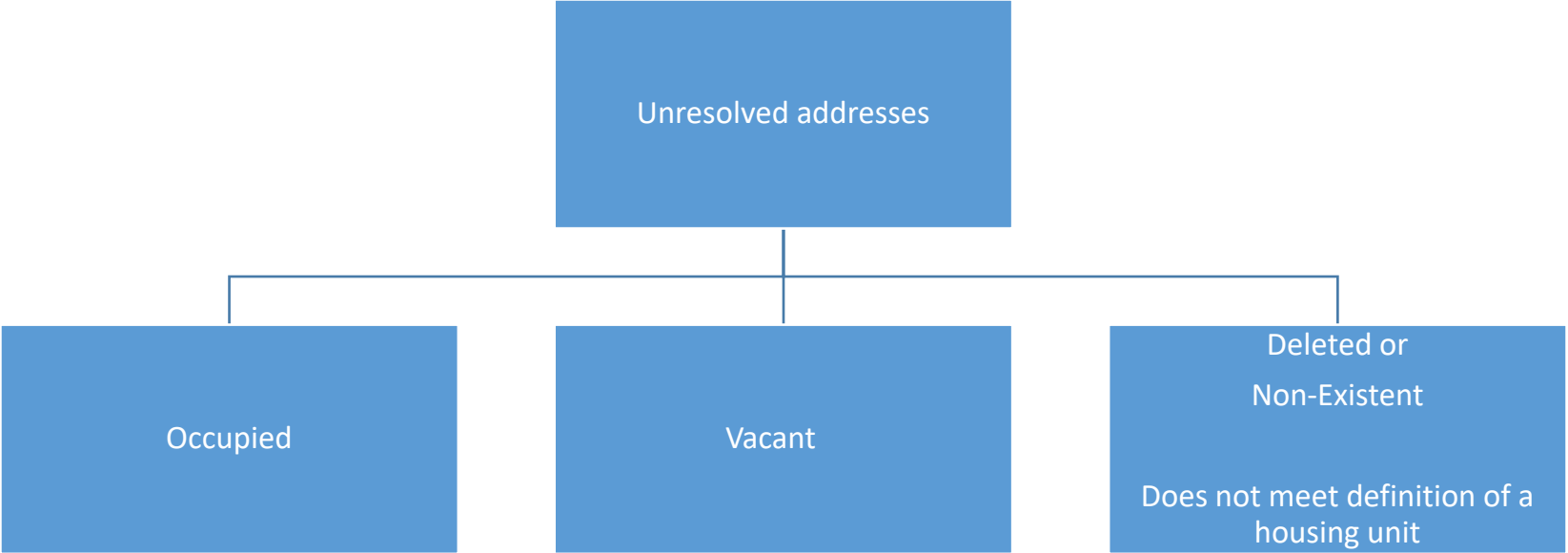
Major initiative area of research and development this decade included:

- Administrative record modeling to support the reduction of contacts in the Nonresponse Followup Operation
- Assigning a census ID to self response without one (NonID operation)
- Self-Response and Nonresponse Followup Quality Assurance

Present preliminary results about the quality of administrative record usage for enumeration and characteristics

- Presentation today will be sharing information available at this time

Administrative Record Enumeration:
Nonresponse Followup (NRFU) Operation Statuses



Quality of Administrative Record Sources

The 2020 Census attempted to use multiple sources for enumeration in the determinations

Administrative Record Occupied multiple sources:

Internal Revenue Service 1040 and 1099/W2 information	People and dependents with income
Center for Medicare and Medicaid Services (CMS) Medicare Enrollment Database	Elderly population
Indian Health Service Patient Database	American Indian and Alaskan Native
Census Bureau Household Composition Key File	Link children to parents

Administrative Record Vacant and Delete/Non-existent sources used in determinations:

- United States Postal Service Postal Carriers indicating Undeliverable-As-Addressed (UAA) reasons for Census mailings around April 1st
- Persons present at address on administrative record or third-party sources

2020 Census Research Using 2010 Census Data

AR Determination by Percent Hispanic Population in Block Group

ACS 5-Year Estimate of Percent of Block Group that is Hispanic	2010 NRFU Addresses (millions)	AR Determination (row percent)		
		AR Vacant (%)	AR Occupied (%)	Full Contacts (%)
0 to 10 percent	31.3	11.6	15.9	72.5
10 to 20 percent	6.8	9.4	15.7	74.9
20 to 30 percent	3.6	8.5	14.4	77.1
30 to 40 percent	2.2	7.6	13.3	79.1
40 to 50 percent	1.6	7.5	12.0	80.5
50+ percent	4.2	4.3	9.6	86.1
Total	49.8	10.1	15.0	74.9

As concentration of Hispanic population increases:

- More full contacts
- Less administrative record usage

Source: March 2017 Census Scientific Advisory Committee Presentation

Quality of Administrative Record Sources for Characteristic Imputation

Past Census or American Community Survey responses for Race, Hispanic origin, age and sex

- 2010 Census used 2000 Census or previous ACS responses for Race and Hispanic origin

Keller (2018) AAPOR Presentation documented quality metrics for other administrative record sources for characteristics:

Among 2010 Census persons reporting their Hispanic origin...

- Those born in a Hispanic country of origin as indicated on the SSA Numident reported being Hispanic 98% of the time
- Those on Census Bureau Best Hispanic origin file agree with 2010 Census report 96% of the time

Among 2010 Census persons reporting their race...

- Those with foreign country of origin as indicated on the SSA Numident reported the same race category 90% of the time
- Those on Census Bureau Best Race file agree with 2010 Census report the same 92% of the time

Among 2010 Census persons reporting their age

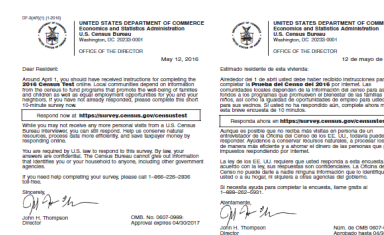
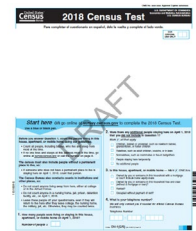
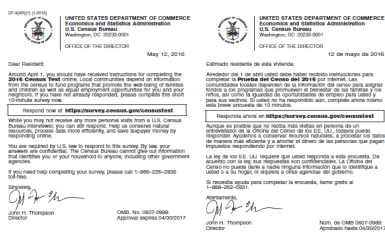
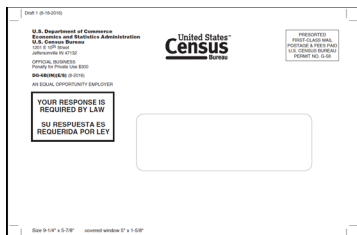
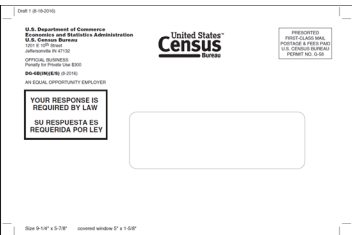
- 98% of persons on SSA Numident agree

SSA is the Social Security Administration

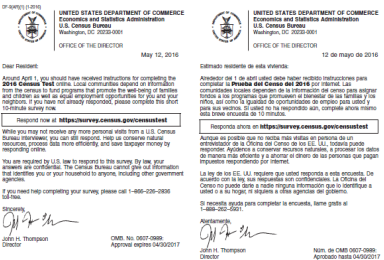
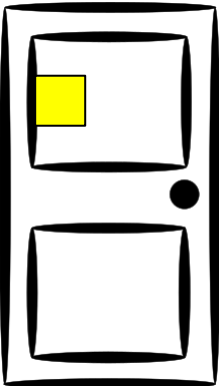
NRFU Contact Strategy and Administrative Record Usage

2020 Census Plan: Using Administrative Records to Enumerate NRFU Housing Units

- #1 Initial letter
- #2 Reminder letter
- #3 Reminder postcard
- #4 Questionnaire
- #5 Not too late postcard



- #6 First visit by enumerator and notice of visit
- #7 Final postcard about one week after visit



Receive a self-response return – we use the respondent provided data

No return received – we use Administrative Records Data



This example is for Self Response TEA1.

2020 Census Plan: Identifying Vacant and Nonexistent Addresses Operational Flow

MARCH 2020

SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3				
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

April 2020

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

June 2020

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

Use administrative records to determine possible vacant and nonexistent address

Address has to have at least one UAA in Census mailings

Send mailing to address in early June about 8 weeks after Census Day

Address receives one field visit
June Mailing undelivered and no sign of occupancy

Field work resolution or self-response

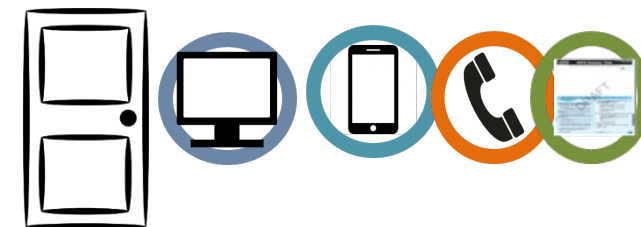
Administrative record vacant

Administrative record nonexistent address

Address receives one field visit
June Mailing delivered or June Mailing undelivered but sign of occupancy

Address has opportunity to self-respond

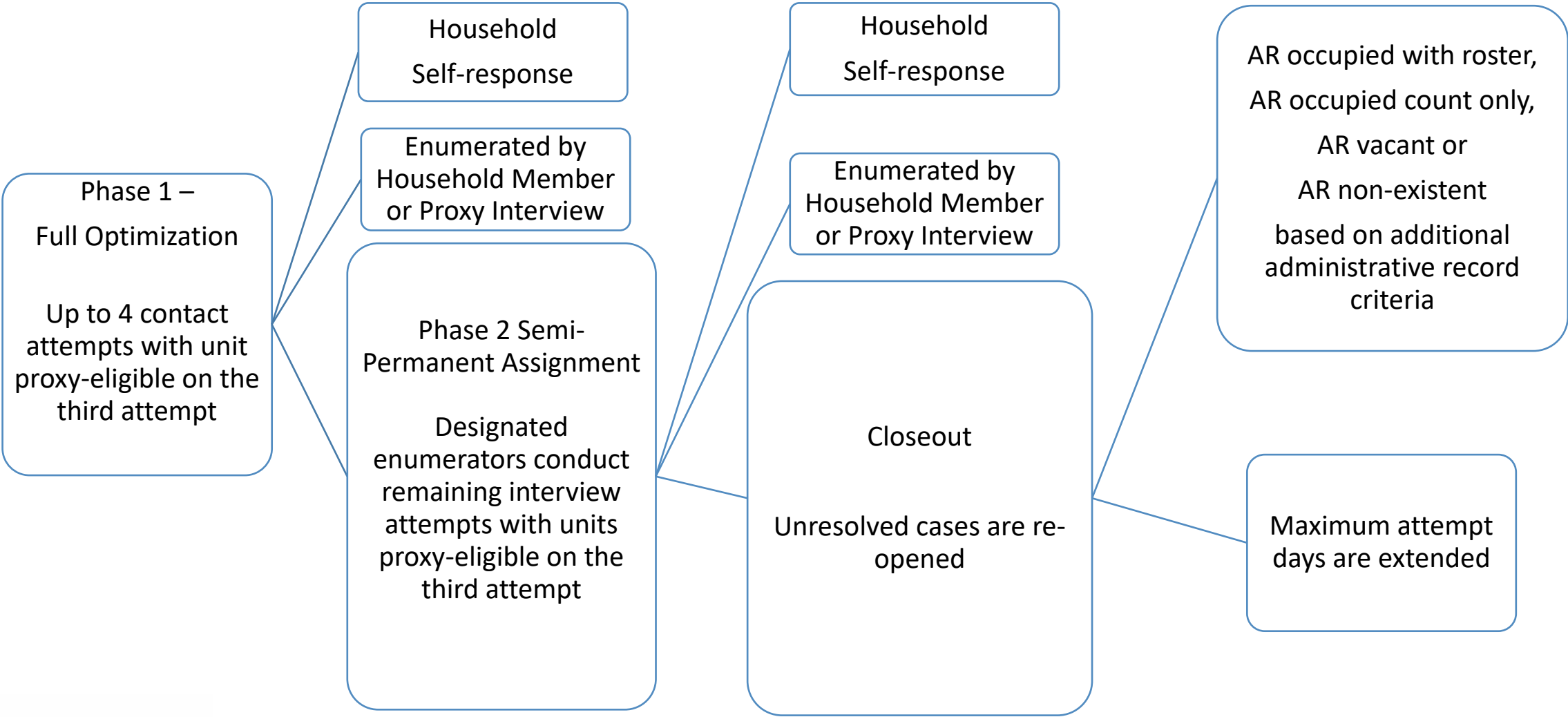
Address receives full NRFU contact strategy



VACANT



NRFU Full Contact Strategy



Covid-19 Implications For NRFU and Administrative Record Usage

Changes to the IRS Tax Filing Deadline and NRFU Operation

Internal Revenue Service (IRS)

- On March 21, 2020, IRS announced that the tax filing deadline was extended from April 15, 2020 to July 15, 2020
 - While there was a delay in the deadline, the Census Bureau continued to receive monthly deliveries of processed 1040 records.

2020 Census

- Start of the NRFU operation delayed until August 9, 2020
 - Soft-launch of NRFU operation started on July 16, 2020 in selected ACOs

Revisions to Administrative Record Modeling in 2020

- To mitigate potential coverage error, changes in May and June modeling resulted in 2.2 million addresses receiving full contacts that would have otherwise been classified as AR vacant and AR delete
- Introduced an occupied model for American Indian Reservations
- Roster determination for AR enumeration if a household did not self-respond was adjusted to allow household members included on later filed tax returns to be included
- Closeout processing expanded to include sole-source determination where we could use population count-only information
- Implementation of the off-campus student initiative to help improve the census results of students whose local school address may have been determined to have been unoccupied.
- Additional administrative record usage for unresolved addresses in four Louisiana Hurricane Parishes
- See Mulry, Mule, Keller and Konicki (2021) Administrative Record Modeling in the 2020 Census for more information

2020 Census

2020 Edit and Characteristic Imputation Overview

Edits

1. During the editing phase, responses are run through a series of checks to:

- a) Detect and correct out-of-range or inconsistent values
- b) Remove invalid or duplicate responses
- c) Convert date of birth values to age values and race/Hispanic origin responses to a numeric code

The best case scenario is that all responses are valid and consistent – we call those “reported” and no imputation is needed. When that doesn’t happen, we use characteristic imputation.

Characteristic Imputation

2. Assignment – responses are missing or inconsistent and information can be determined from:

- a) Other responses provided for that same person
- b) Previous Census or Administrative Records (AR)



3. Allocation – responses are missing or inconsistent and information can be determined from:

- a) Responses provided for other persons in the household
- b) Similar nearby households using “nearest neighbor hot decks”

Preliminary Quality Measures for Administrative Record Usages in the 2020 Census

Preliminary results for AR Modeling and Enumeration:

- How often were different types of AR occupied outcomes used?
 - One Visit with rosters, Closeout with rosters, Closeout with population count only
- What is the count comparison of AR rosters to Self-Response and NRFU HH enumerations?
- What was the availability of characteristics for AR Enumerations?

Preliminary results for usage of administrative records in characteristic imputation:

- How often was past Census or administrative record information used for select characteristics?
- How did reported race and Hispanic origin in the 2020 Census agree with past Census or administrative record information?

Preliminary Measures For Administrative Record Modeling and Enumeration

2020 AR Enumerations by Roster Quality Category

4.59 percent of 151.8 million addresses nationwide have been completed using high-quality administrative records

- 3.20 percent occupied
- 1.15 percent vacant
- 0.24 percent delete

April 2021 Release 1: CBDRB-FY21-DSSD007-0012

- For the 3.20 percent AR occupied, how often were different types used?

AR Roster Quality Category	Count	Percent of 4.859 million AR Occupied addresses	Percent of 151.8 million Total addresses
One-visit with rosters	4,468,000	92.0%	2.9%
Closeout with rosters	186,000	3.8%	0.1%
Closeout Household size only	188,000	3.9%	0.1%
Other	16,500	0.3%	<0.1%
Total	4,859,000	100.0%	3.2%

CBDRB-FY22-172 Due to rounding percentages may not add to 100 percent

How does the AR enumeration counts compare to Census Responses?

- Took advantage that AR modeling was run on the entire census enumeration universe during 2020 production
 - Focus on the One Visit multiple source (2.2 people per HH)
 - Built rosters that would have been used for enumeration
 - Compare those roster counts to census responses
- Census Response Types
 - Self Response
 - NRFU Householder
- Timing of Census Response checked into the control system
 - Before July 30th
 - After July 30th

Count Comparisons of Administrative Enumerations vs Census Responses

	Household size difference (=Census HH size – AR HH size)								Total count
	AR Higher			Same	Census Higher				
	<=-3	-2	-1	0	1	2	3>=	Total	
One-visit multiple source (2.2 people per HH)									
Self-Response									
Resolved by July 30 th	0.2%	0.5%	6.1%	83.8%	7.4%	1.3%	0.6%	100.0%	44,540,000
Unresolved by July 30 th	0.6%	1.0%	8.3%	74.4%	11.7%	2.6%	1.4%	100.0%	1,957,000
NRFU HH member	1.4%	2.1%	10.7%	66.3%	13.7%	3.6%	2.2%	100.0%	2,083,000
CBDRB-FY22-172									

Administrative records had a count agreement of 83.8 percent when compared to self-responses before July 30th

Count agreement decreases as time gets further away from Census Day and mode changes

Census responses have potential recall bias as it gets further from Census Day and mode effects due to interviewer vs self-response

Availability of characteristics for assignment for 2020 AR Enumerations

Characteristic	Census 2020
Age	96.4%
Sex	96.5%
Race or Hispanic Origin	83.6%

CBDRB-FY22-172

If the characteristics were not available from past census or administrative record sources then the information would have to imputed

Preliminary Measures
For
Past Census and Administrative Record Usage for
Demographic Characteristic Items

National Usage of Past Census and Administrative Records during Characteristic Imputation Processing

Preliminary usage results for 331.4 million census people:

	Percent of 331.4 million assigned from	
Characteristic	Past Census or ACS response	Other Administrative Records
Race	3.2%	1.5%
Hispanic Origin	3.2%	0.9%
Age	2.6%	1.4%
Sex	2.6%	1.2%

CBDRB-FY22-172

Preliminary Agreement of Race and Hispanic Origin responses to Past Census and Administrative Record assignments

Preliminary Race and Hispanic Origin Analysis:

14 categories of Race and Hispanic origin assignment:

Non-Hispanic	Hispanic
White Alone	White Alone
Black Alone	Black Alone
American Indian and Alaska Native Alone	American Indian and Alaska Native Alone
Asian Alone	Asian Alone
Native Hawaiian and Pacific Island Alone	Native Hawaiian and Pacific Island Alone
Some Other Race Alone	Some Other Race Alone
Multiracial	Multiracial

2020 Census reports and AR assignment available and agree:

- 89% of the 14 Race/Hispanic Origin assignment is the same.

Summary

- Multiple sources were used for enumeration and characteristic assignment purposes
- Administrative record enumerations and characteristics were only used if a self-response or resolution for an address was not obtained.
- 92 percent of AR enumerations received one visit in NRFU with rosters and available characteristics based on multiple sources indicating the family lived at the address
- Only 3.9 percent of the AR enumerations were from the closeout operation where only the population count was used.
- The count agreement between AR enumerations and Census self-response received before July 30th was 83.8 percent
- Shared preliminary result how of usage of administrative records for characteristics including usage of past census or ACS information
- Preliminary results show 89 percent agreement when comparing census responses to available administrative record assignments

Questions for the Committee

As the Census Bureau starts planning for the 2030 Census:

- How do you think administrative records should be used in the next census?
- Are there sources that you can suggest the Census Bureau pursue to improve the coverage of historically undercounted populations?